AFHRL-TR-90-3

AIR FORCE

HUMAN RESOURCES

AIR FORCE OFFICER QUALIFYING TEST (AFOQT): DEVELOPMENT OF QUICK SCORE COMPOSITES FOR FORMS P1 AND P2

Todd C. Sperl
Malcolm James Ree

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

June 1990

Interim Technical Report for Period November 1987 - January 1990

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

This report has been reviewed and is approved for publication.

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1990

3. REPORT TYPE AND DATES COVERED
Interim - November 1987 - January 1990

4. TITLE AND SUBTITLE
Air Force Officer Qualifying Test (AFOQT):
Development of Quick Score Composites for Forms P1 and P2

5. FUNDING NUMBERS
PE - 62703F
PR - 7719
TA - 18
WU - 47

6. AUTHOR(S)
Todd C. Sperl
Malcolm James Ree

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8. PERFORMING ORGANIZATION REPORT NUMBER
AFHRL-TR-90-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This work was accomplished under SC Study 9495.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this effort was to develop an instrument to predict success on the Air Force Officer Qualifying Test (AFOQT) Forms P. This device, referred to as Quick Score Composites (QSC), will aid in the identification of individuals likely to pass AFOQT requirements in applying for a commission in the Air Force. The subjects were 6,192 applicants administered AFOQT Forms P1 and P2. Two item selection methods (point-biserial and random) were compared for accuracy, reliability, and control of differential score prediction for gender and ethnic groups. Results showed that QSC scores based on random item selection within subtest are effective predictors of the AFOQT composites. This was indicated by their comparable reliability to that of the other method, high positive correlations with the composites they represent, and introduction of less gender and ethnic bias than that introduced by the point-biserial method.

14. SUBJECT TERMS
Air Force Officer Qualifying Test (AFOQT)
item selection
prescreening tests
quick score composites
short forms
test bias

15. NUMBER OF PAGES
40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

SUMMARY

This report describes the development of short-form screening composites called Quick Score Composites (QSCs), designed to assist Air Force test administrators and recruiters in obtaining quick and accurate estimates of applicants' aptitude scores on the Air Force Officer Qualifying Test (AFOQT), Forms P. AFOQT scoring is centralized, and the turnaround time for reporting official score results to applicants and recruiters in the field is about 1 to 2 weeks. The delay slows processing of applicants and may have a negative impact on recruiting capability since potential recruits may seek other employment opportunities. The Quick Score procedure is needed to help recruiters prescreen applicants and to expedite the processing of those candidates with a high likelihood of meeting aptitude requirements for commissioning. The QSCs consist of abbreviated versions of the five composites on the AFOQT Forms P: Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative. Two item selection methods (point-biserial and random) were compared for accuracy, reliability, and existence of differential score prediction for gender and ethnic groups. It was concluded that QSC scores based on random item selection within subtest are effective predictors of the full-length scales. This was indicated by their comparable reliability to that of the point-biserial method, high positive correlations with the scales they represent, and introduction of less gender or ethnic bias than that introduced by the other method.

PREFACE

The Air Force Human Resources Laboratory (AFHRL) is tasked as the test development agency for the Air Force Officer Qualifying Test (AFOQT) by Air Force Regulation 35-8, Air Force Military Personnel Testing System. The current research and development (R&D) effort was undertaken as part of AFHRL's responsibility to develop, revise, and conduct research in support of the AFOQT. This research was completed under 771918, Selection and Classification Technologies, which is part of a larger effort in Force Acquisition and Distribution Systems. It was subsumed under work unit number 77191847, Development and Validation of Civilian and Nonrated Officer Selection Methodologies.

The authors would like to thank their colleagues in the Manpower and Personnel Division for assistance in this effort. Ms. Jacobina Skinner provided generous expert advice on a variety of technical and editorial issues. A number of other personnel contributed to the successful accomplishment of this study. Specifically we extend our appreciation to Drs. William E. Alley, Lonnie Valentine, Jr., and Kurt W. Steuck.

The authors acknowledge with considerable gratitude the assistance of Ms. Doris E. Black, Mr. James L. Friemann, A1C Robert C. Stine, SRA William J. Myers, and SRA James E. Bloch of the Information Sciences Division, AFHRL. Their efforts were instrumental to the successful accomplishment of the data analyses for this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Tables (Concluded)

AIR FORCE OFFICER QUALIFYING TEST (AFOQT):
DEVELOPMENT OF QUICK SCORE COMPOSITES FOR FORMS P1 AND P2

I. INTRODUCTION

Since 1953, the Air Force Officer Qualifying Test (AFOQT), a multiple-aptitude test battery, has been part of the selection process for officer commissioning training. Training programs are conducted by the Officer Training School (OTS) at Lackland AFB, Texas, and by the Air Force Reserve Officer Training Corps (AFROTC) on college campuses. Test results are also considered in selecting recipients of AFROTC scholarships and in classifying commissioned officers into pilot and navigator specialties (Rogers, Roach, & Short, 1986).

Short-form screening composites have been associated with the AFOQT since its implementation 37 years ago. During this time, several prescreening devices with different characteristics have been developed. Despite their differing features, all have shared the common goal of reducing time and costs associated with applicant testing and processing.

Early screening composites were derived from separate tests containing items of the same type and content as those in the corresponding subtests on the full-length AFOQT (Miller, 1966, 1968; Valentine, 1961).

The Air Force Precommissioning Screening Test (AFPST), the first short form associated with the AFOQT testing program, was a separate test used to screen applicants for the Air Force Academy Preparatory School and navigator training (Valentine, 1961). The AFPST was a continuation of an effort begun in 1949, to develop a short test instrument for the screening of aircrew applicants prior to administration of a longer and more sophisticated test battery. A more extensive description of past efforts in the development of screening tests is presented in Valentine and Creager (1961).

The 10 subtests of the AFPST were arranged into five parts or composites: Verbal, Quantitative, General Science, Mechanical, and Scale Reading. The AFPST consisted of items that were similar in content and difficulty level to items found in the AFOQT Form G. Most of the items selected for the AFPST were derived from a previous AFOQT, specifically Form F (Valentine, 1961).

A revised AFPST, named the AFROTC Pre-Enrollment Test (PET-66), was implemented along with the AFOQT-66 in the summer of 1965 (Miller, 1966). It was intended that the PET-66 would screen freshman applicants for the AFROTC program. The test was first implemented at AFROTC detachments during the summer and fall of 1965 (Miller, 1966).

Because the PET-66 was to focus on the screening of freshmen AFROTC applicants, the specific Navigator-Technical subtests found in the AFPST were deleted. The remaining Verbal and Quantitative subtests were arranged to form three composites: Verbal, Quantitative, and Total (Verbal + Quantitative). Each subtest contained 30 items which were selected from a previous AFOQT form. The criteria used for the selection of items were similar to those used in the development of the AFPST. Items were selected on the basis of high internal consistency (correlation with total test score) and composite difficulty levels. The testing time was slightly over 1 hour, with single total score being recorded. The properties of the PET-66 have been reported in detail elsewhere (Miller 1966).

The PET-68, a revision of the PET-66, was implemented in 1967, to coincide with the administration of the AFOQT-68, which met the requirement from Headquarters USAF to establish a 2-year revision cycle for the AFOQT (Miller, 1968). The PET-68 continued as a short device for screening freshmen applicants for the AFROTC program. Because of scheduling difficulties with testing time, the PET-68 was made shorter than the PET-66. To accomplish this, the number of items was reduced to include only 20 items from both the Verbal and Quantitative portions of the test. Once again a single score based on number of correct responses was obtained, with test administration time being reduced to under 1 hour. Further details on the properties of the PET-68 can be found in Miller (1968). The PET-68 was discontinued in April 1969, leaving recruiters without a prescreening device for officer applicants until the implementation of AFOQT Form O.

The implementation of Form O introduced centralized scoring of answer sheets to the AFOQT testing program (Rogers, 1985). The time delay (1 to 2 weeks) for reporting official score results associated with the new centralized scoring process prompted the Air Force to reinstate the use of short-form screening tests. The Air Force Human Resources Laboratory (AFHRL) was tasked to develop screening composites for Form O, called Officer Screening Composites (OSCs).

The use of a separate short test raised concerns about retest effects and potential for compromising the full AFOQT. Form O prescreening procedures addressed these issues by basing the screening composites on a subset of the items in the full AFOQT (Rogers, 1983). This strategy eliminated possible retest effects and reduced the possibility of test compromise.

Form O OSCs were scored by test administrators using template scoring keys which, when placed over the Form O answer sheet, identified the Form O items to be scored for the OSCs. This process of scoring the OSCs took roughly 1 hour and could be performed at any testing location. The properties of these OSCs have been reported in detail elsewhere (Rogers, 1985; Wegner & Short, 1986).

A long-term problem associated with AFOQT short-form screening composites concerns the representativeness of screening test coverage. For example, the Officer Screening Composites used with the AFOQT Form O from 1982 to 1988 omitted several subtests. Any overrepresentation and underrepresentation of items from specific subtests would change the taxonomic structure and could possibly produce unwanted gender and ethnic bias in the prediction of full AFOQT scores from the prescreening composites.

The present effort was designed to improve prescreening procedures for AFOQT Form P, the operational form which replaced Form O in 1988. For the first time, parallel versions of the AFOQT--Forms P1 and P2--were developed. In view of this innovation in officer testing practice and some item content changes from Form O to Form P, Form O Officer Screening Composites were not appropriate for use with Form P. The purpose of the present effort, therefore, was to develop accurate and reliable screening composites to aid recruiters in identifying those applicants for OTS and AFROTC most likely to succeed on AFOQT Form P.

The approach taken to develop Form P's prescreening procedure, called Quick Score (QS), compared two methods of selecting item subsets from the full AFOQT: (a) items with the highest point-biserials, and (b) items selected randomly. These methods were judged most promising among those described in the literature on short-form scoring (Bryson, 1972; Goh, 1979; Henrysson, 1979; Levy, 1968; Miller, 1966, 1968; Mumpower, 1964; Rogers, 1983; Valentine, 1961; Wegner & Short, 1986).

Goh (1979) performed a similar study in which he compared two item selection techniques for accuracy; items were selected both empirically and randomly. He observed that from a psychometric viewpoint, it would be more desirable to select test items for the short form on a systematic empirical basis. He concluded that selecting items with the highest index of discrimination and an item difficulty close to .50 was preferable to random selection. Further, he noted that selecting items psychometrically, rather than randomly, should increase the probability that a short-form screening test will correlate highly with the full-scale test.

However, Goh used the Yudin (1966) "random" item selection method, which may not have been truly random in that it selects items from subtests on the basis of every Xth item and allows X to vary from one subtest to another. A true random selection of items should produce short-form screening tests which are more representative not only of the range of item discriminative power but also of the difficulty and content of the items in the full-length test (Neter, Wasserman, & Whitmore, 1988).

The current study compared the two methods (point-biserial versus random) for accuracy, reliability, and control of differential score prediction for gender and ethnic groups.

II. METHOD

Subjects

Subjects were 6,192 officer applicants administered the AFOQT. Forms P1 and P2 were administered to 3,216 and 2,976 subjects, respectively. As shown in Table 1, a majority of the subjects were white males

(80%) and, on average, 22 years of age. All had completed at least a secondary education program by diploma or certificate of equivalency. About 36% held college degrees (baccalaureate or higher).

Table 1. AFOQT Forms P1 and P2 Demographic Percentages

Gender	P1	P2	Ethnic	P1	P2
Male	80.9	79.7	Caucasian	80.2	80.0
Female	19.0	20.1	Black	10.9	11.2
			Hispanic	3.8	3.8
			Asian	4.0	3.5
			AI/AN[a]	.8	1.3
Degree	P1	P2			
High School	53.1	55.3			
Bachelor	36.3	35.2			
Associate	8.0	6.6	Age	P1	P2
Master	1.9	2.2			
Doctoral	.1	.2	Average	22.3	22.0

[a]American Indian/Alaskan Native.

The total number of subjects used in gender and ethnicity analyses was reduced from the total sample due to a small number of examinees in the American Indian/Alaskan Native group. These examinees were excluded from the linear models analyses because there were so few such examinees that statistical tests of their data would lack power. Examinees in four other groups (Caucasian, Black, Hispanic, and Asian) were included in the gender and ethnicity analyses.

Instrument

The AFOQT Form P is a paper-and-pencil instrument with multiple-choice test items designed for group administration under standardized conditions. As shown in Table 2, the test contains 380 items, grouped into 16 subtests of 15 to 40 items each. The subtests are aggregated into five composites: Pilot (P), Navigator-Technical (N-T), Academic Aptitude (AA), Verbal (V), and Quantitative (Q).

Prior analyses of Forms P1 and P2 indicate that these forms are generally similar to each other and are comparable to Form O in terms of content, item difficulty, and discriminative power (Steuck, Watson, & Skinner, 1988). A further description of Form P test construction can be found elsewhere (Berger, Gupta, Berger, & Skinner, 1988).

Procedure

Subjects were administered Form P of the AFOQT between June and October 1987, at Military Entrance Processing Stations (MEPS), AFROTC detachments located on college and university campuses, and at Consolidated Base Personnel Offices (CBPOs) on Air Force installations. Examinees' responses were collected on machine-scorable answer sheets.

Item Selection

Two item selection methods were investigated. The first method identified items based on the highest item-subtest point-biserial correlations. The second method identified items using random selection procedures.

Table 2. **Composition of AFOQT Form P Aptitude Composites**

Subtest	Number of items	Composite P	N-T	AA	V	Q
Verbal Analogies	25	X		X	X	
Arithmetic Reasoning	25		X	X		X
Reading Comprehension	25			X	X	
Data Interpretation	25		X	X		X
Word Knowledge	25			X	X	
Math Knowledge	25		X	X		X
Mechanical Comprehension	20	X	X			
Electrical Maze	20	X	X			
Scale Reading	40	X	X			
Instrument Comprehension	20	X				
Block Counting	20	X	X			
Table Reading	40	X	X			
Aviation Information	20	X				
Rotated Blocks	15		X			
General Science	20		X			
Hidden Figures	15		X			
Total	380					

To ensure equal representation of all AFOQT subtests in the Quick Score Composites (QSCs), items from all subtests were used in both item selection methods. Further, to guard against overrepresentation or underrepresentation of items from specific subtests, an equal percentage of items was selected from each subtest. Review of past AFOQT screening composites (Miller, 1966, 1968; Rogers, 1983; Valentine, 1961) led to the decision to select 27% of the original items, about the same percentage that was selected for Form O Officer Screening Composites. Table 3 indicates the number of items selected from each subtest using both the point-biserial and random methods. The "Number of Items Selected per QSC" columns indicate the number of items selected from each subtest/composite in the full AFOQT.

Variables

The variables used in the analyses were (a) AFOQT Form P raw scores on the full Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative composites; (b) raw scores on the Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative point-biserial QSCs; and (c) raw scores on the Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative random QSCs. Raw scores were computed as the number of correct answers to items contained in the specific scale.

Analysis

For analytic purposes, Form P1 and Form P2 samples were randomly divided into validation and "cross-validation" groups. Form P1's 3,216 subjects were randomly divided into two samples of 1,608 subjects each; the 2,976 Form P2 subjects were randomly divided into two samples of 1,488 subjects each.

The initial analyses and item selections were done on the validation groups. The cross-validation groups were used to replicate and cross-validate the statistical properties of scores and relationships found in the validation group.

Table 3. **Item Content of AFOQT Forms P1 and P2 Quick Score Composites**

Subtest	Items scored in full AFOQT	Number of items selected per QSC P	N-T	AA	V	Q
Verbal Analogies	25	7		7	7	
Arithmetic Reasoning	25		7	7		7
Reading Comprehension	25			7	7	
Data Interpretation	25		7	7		7
Word Knowledge	25			7	7	
Math Knowledge	25		7	7		7
Mechanical Comprehension	20	5	5			
Electrical Maze	20	5	5			
Scale Reading	40	11	11			
Instrument Comprehension	20	5				
Block Counting	20	5	5			
Table Reading	40	11	11			
Aviation Information	19	5				
Rotated Blocks	15		4			
General Science	19		5			
Hidden Figures	15		4			
Total	378[a]	54	71	42	21	21
Full AFOQT Form P		204	264	150	75	75

[a]Of the 380 AFOQT items administered to the subjects, two items were deleted from scoring for not having clearly identifiable answers. The same items were excluded from the present analyses.

After development of the validation and cross-validation groups for both Forms P1 and P2, group characteristics were inspected for comparability. Gender, ethnicity, type of degree earned, and age were compared to those of the overall sample to assure that the validation and cross-validation groups were representative of the total sample.

Several analyses were performed to determine which QSC (point-biserial or random) best represented its full AFOQT composite. For each QSC, its average item difficulty level was compared to that of its respective full AFOQT composite. Other descriptive statistics computed for the random and point-biserial QSCs included: mean, standard deviation, kurtosis, skewness, minimum and maximum. To assess the accuracy of the two item selection methods, Pearson product-moment correlations were computed between random and point-biserial QSC raw scores and the raw scores for their respective full AFOQT composites. Reliability analyses were conducted on each QSC using procedures developed by Cronbach (1951) and Wherry and Gaylord (1943).

Linear models analyses were performed to test for gender and ethnicity effects in the QSCs. Linear models analysis is a regression-based technique in which a full model is compared with one or more restricted models through the use of F-tests. Each comparison between the full and restricted models is evaluated using the statistic and associated probability value:

$$F = \frac{(R_f^2 - R_r^2) / df_1}{(1 - R_f^2) / df_2}$$

where

R_f^2 = Squared multiple correlation - full model

R_r^2 = Squared multiple correlation - restricted model

df_1 = Number of independent predictor variables in the full model minus the number of independent predictor variables in the restricted model

df_2 = Total number of observations minus the number of independent predictor variables in the full model.

The computational procedure used to determine whether bias is present first tests to determine the need for squared terms (X^2) to represent the data. Then the null hypothesis of no difference (equal) between slopes and intercepts (i.e., whether the regression lines are identical) is tested. If this null hypothesis is rejected, further tests must be made to determine whether the bias that is present is one of slope or of intercept. A complete explanation of this procedure may be found in Ward and Jennings (1973). Tables showing the full and restricted models for analysis of ethnic and gender bias, as well as decision tree diagrams, are provided in Appendix A.

III. RESULTS AND DISCUSSION

Composite Difficulty Levels

Analyses were conducted to obtain the average item difficulty levels for the five composites. The average composite difficulty levels did not vary between the QSCs and full AFOQT composites by more than .08, and differences were typically much smaller (see Tables 4 and 5). In both Forms P1 and P2, a majority of the random QSC mean composite difficulty levels were slightly higher than those of the full AFOQT Form P1 and the point-biserial mean composite difficulty levels. Tables B-1 and B-2 indicate that this same pattern held true when replicated on a comparable sample.

Table 4. **Mean Composite Difficulty Levels of Form P1 for Full, Point-Biserial, and Random Composites**

Composite	AFOQT Form P1	Pbis QSC	Diff[a]	Random QSC	Diff[a]
Pilot	.57	.53	.04	.60	-.03
Navigator-Technical	.59	.58	.01	.63	-.04
Academic Aptitude	.63	.58	.05	.66	-.03
Verbal	.62	.59	.03	.66	-.04
Quantitative	.64	.56	.08	.66	-.02

[a]Indicates difference between full length AFOQT and item selection method.

Table 5. **Mean Composite Difficulty Levels of Form P2 for Full, Point-Biserial, and Random Composites**

Composite	AFOQT Form P2	Pbis QSC	Diff[a]	Random QSC	Diff[a]
Pilot	.58	.54	.04	.60	-.02
Navigator-Technical	.58	.53	.05	.60	-.02
Academic Aptitude	.64	.58	.08	.63	.01
Verbal	.65	.59	.06	.65	.00
Quantitative	.63	.57	.06	.62	.01

[a]Indicates difference between full length AFOQT and item selection method.

It should be noted that point-biserial correlations reach their maximum only when item difficulty is exactly .50. Consequently, items at or near this value tend to be preferred by the point-biserial item selection method. This accounts for the lower mean difficulty level of the point-biserial QSCs. The random item selection method produces QSCs that contain both easier and more difficult items. Nowhere in the literature on screening tests is this artifact discussed, despite the well-known nature of the point-biserial correlation coefficient.

Summary Statistics of QSCs

Results showed that Forms P1 and P2 random QSC mean composite scores were, on average, significantly higher ($p < .01$) than point-biserial QSC mean composite scores (see Tables 6 through 9). For example, Form P1 point-biserial QSC mean composite raw scores ranged from 11.70 to 37.46 (see Table 6) while Form P1 random QSCs had significantly higher ($p < .01$) mean raw scores, ranging from 13.86 to 44.57 (see Table 7). This same pattern was found when the analysis was replicated on a comparable sample (see Tables B-3 through B-6).

Table 6. **Summary Statistics of AFOQT Form P1 Point-Biserial Quick Score Composites**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	28.50	11.08	-.78	-.09	1	53
Navigator-Technical	37.46	14.98	-.84	-.10	2	70
Academic Aptitude	24.18	9.56	-.85	-.22	0	42
Verbal	12.48	5.10	-.84	-.27	0	21
Quantitative	11.70	5.63	-1.04	-.10	0	21

Table 7. **Summary Statistics of AFOQT Form P1 Random Quick Score Composites**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	32.55	7.48	-.38	-.23	7	49
Navigator-Technical	44.57	10.81	-.44	-.32	11	69
Academic Aptitude	27.82	7.25	-.45	-.42	6	42
Verbal	14.00	3.76	-.24	-.51	2	21
Quantitative	13.86	4.50	-.75	-.33	2	21

Table 8. **Summary Statistics of AFOQT Form P2 Point-Biserial Quick Score Composites**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	29.19	10.92	-.71	-.22	1	53
Navigator-Technical	37.56	14.65	-.76	-.19	2	68
Academic Aptitude	24.28	10.04	-.92	-.20	1	42
Verbal	12.30	5.43	-.96	-.17	0	21
Quantitative	11.98	5.74	-1.12	-.16	0	21

Table 9. **Summary Statistics of AFOQT Form P2 Random Quick Score Composites**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	32.47	8.43	-.49	-.28	8	51
Navigator-Technical	42.89	10.73	-.40	-.34	11	67
Academic Aptitude	26.61	7.76	-.61	-.30	6	42
Verbal	13.58	4.38	-.67	-.30	2	21
Quantitative	13.02	4.43	-.78	-.25	1	21

Scores for point-biserial QSCs were generally more variable than the random QSC scores (see Tables 6 through 9). The standard deviation of each QSC raw score is a good measure of the absolute variability. However, the Coefficient of Variation (CV), a measure of relative variability, compares the QSC raw score variability to the full AFOQT composite raw score variability. The Coefficient of Variation is the ratio of the standard deviation to the mean expressed as a percentage (Neter et al., 1988). Results indicated that random QSCs were of relatively the same variability as the full AFOQT composites, whereas the point-biserial QSCs consistently showed greater relative variability (see Tables 10 and 11). Replication of these analyses on a comparable sample can be found in Appendix B (Tables B-7 and B-8).

Table 10. **Coefficient of Variation of AFOQT Form P1 Full, Point-Biserial, and Random Composites**

Composite	Full	Pbis QSC	Random QSC
Pilot	24.01	38.88	22.98
Navigator-Technical	24.40	39.99	24.25
Academic Aptitude	26.87	39.54	26.06
Verbal	29.03	40.87	26.86
Quantitative	30.28	48.12	32.47

Table 11. **Coefficient of Variation of AFOQT Form P2 Full, Point-Biserial, and Random Composites**

Composite	Full	Pbis QSC	Random QSC
Pilot	23.61	37.41	25.96
Navigator-Technical	24.31	39.00	25.02
Academic Aptitude	27.26	41.35	29.16
Verbal	28.71	44.15	32.25
Quantitative	31.53	47.91	34.02

QSC Reliabilities

Subtest reliabilities for both QSCs were determined by coefficient alpha (Cronbach, 1951). The total score reliability for each QSC was obtained by treating the subtests as components of a composite and applying the Wherry and Gaylord (1943) formula. The reliabilities for Form P1 and P2 QSCs are reported in Tables 12 and 13 (see Tables B-9 and B-10 for replication results).

Table 12. **AFOQT Form P1 Reliabilities for Full, Point-Biserial, and Random Composites**

	Reliability		
Composite	Full	Pbis	Random
Pilot	.958	.931	.857
Navigator-Technical	.971	.950	.909
Academic Aptitude	.961	.926	.886
Verbal	.931	.867	.792
Quantitative	.941	.896	.845

Reliabilities for both the random and point-biserial QSCs were lower than the reliabilities found for the total AFOQT. This is as expected due to the deletion of items, which decreases the reliability. Results indicate that the point-biserial QSCs were slightly more reliable than the random QSCs. This difference was probably

due to the nature of the item selection method. The point-biserial method selected and ranked items by their highest item-total subtest score correlation, providing an opportunity to select highly reliable items.

Table 13. **AFOQT Form P2 Reliabilities for Full, Point-Biserial, and Random Composites**

	Reliability		
Composite	**Full**	**Pbis**	**Random**
Pilot	.957	.932	.880
Navigator-Technical	.969	.948	.908
Academic Aptitude	.963	.935	.897
Verbal	.937	.888	.837
Quantitative	.943	.900	.830

QSC Correlations with Full AFOQT

Table 14 shows the correlations between the full AFOQT composites and the Quick Score Composites. All correlations between QSCs and AFOQT composites are positive and statistically significant at the $p < .01$ level. The high correlations can be accounted for, in part, by the overlap of items.

Table 14. **Correlations Between Full AFOQT Composites and Corresponding Point-Biserial and Random QSCs on Forms P1 and P2**

	Form P1		Form P2	
Composite	**Pbis**	**Random**	**Pbis**	**Random**
Pilot	.940	.942	.937	.956
Navigator-Technical	.955	.964	.956	.964
Academic Aptitude	.957	.950	.960	.958
Verbal	.942	.905	.939	.934
Quantitative	.938	.941	.948	.935

Note. All correlations are significant at $p < .01$ level.

In analyzing Form P1, results showed that correlations between the five point-biserial QSCs and their corresponding full AFOQT scores ranged from .938 to .957. This indicates that about 88% to 92% of the variance in AFOQT Form P1 composite scores can be accounted for by the point-biserial QSC scores. Correlations between random QSCs and their full AFOQT composites varied from .905 to .964, indicating that roughly 83% to 92% of the variance in the AFOQT Form P1 composite scores can be accounted for by the random QSC scores. Results indicated that the same pattern existed for Form P2 QSCs and AFOQT correlations. Correlations ranged from .937 to .960 for point-biserial QSCs and from .934 to .964 for random QSCs.

Tables B-11 and B-12 show the correlations for the full AFOQT and its appropriate QSC for the validation and cross-validation (C-V) samples and the difference between these correlations. The raw score weights from the validation samples were applied to the cross-validation samples and the squared correlations were checked for shrinkage (Allen & Yen, 1979; Mosier, 1951). (See Tables B-13 through B-14.) Statistical tests between the validation and cross-validation correlations were significant ($p < .01$); however, the magnitude of the difference in expected scores was trivial and therefore not of practical significance.

Gender and Ethnicity Analysis

To evaluate gender and ethnic bias in the Quick Score Composites, a series of "step-down" linear models analyses (Lautenschlager & Mendoza, 1986; Ward & Jennings, 1973) were performed on the validation

samples of Forms P1 and P2. This procedure provides the opportunity for testing specific hypotheses about the influence of various predictor variables while holding constant the effects attributable to the remaining variables. The tests were conducted by comparing the errors of prediction associated with a given set of variables (full model) with the errors associated with a reduced set (restricted model) after adjustment for the appropriate degrees of freedom.

The starting full model and the various restricted models for the gender bias analysis contained variables as specified in Table A-1 in Appendix A. The starting full model and restricted models for the ethnic bias analysis can be found in Table A-2 in Appendix A. Table A-3 summarizes the model specifications for the expected relationship between full AFOQT scores and QSC scores in terms of functional form (linear or curvilinear) and between-group effects (interaction, parallel, or no difference). Statistical comparisons between the models were performed sequentially through the network described in Figure A-1 in Appendix A until the most appropriate model was found.

First an initial overall test for linearity was conducted (Model 1 vs. Model 2), followed by a test for a *common slope (Model 1 vs. Model 3 or Model 2 vs. Model 4), generally referred to as a test for no interaction.* If the test for common slope was found to be non-significant, tests for a common intercept (Model 3 vs. Model 5 or Model 4 vs. Model 6) were conducted.

For the purpose of this study, gender effects in the QSCs were considered to exist when the relationship between Quick Score Composites and AFOQT composites differed between genders. Moreover, two different types of bias, intercept and slope bias, were distinguished. When regression lines are parallel but the intercepts are different, "intercept bias" is said to exist; i.e., raw scores on the full AFOQT composite for Subgroup 1 and Subgroup 2 are expected to differ by a constant amount over the entire QSC (see Figure 1A). If there is no consistent underprediction or overprediction for groups, then the test/procedure is considered unbiased (see Cole, 1973).

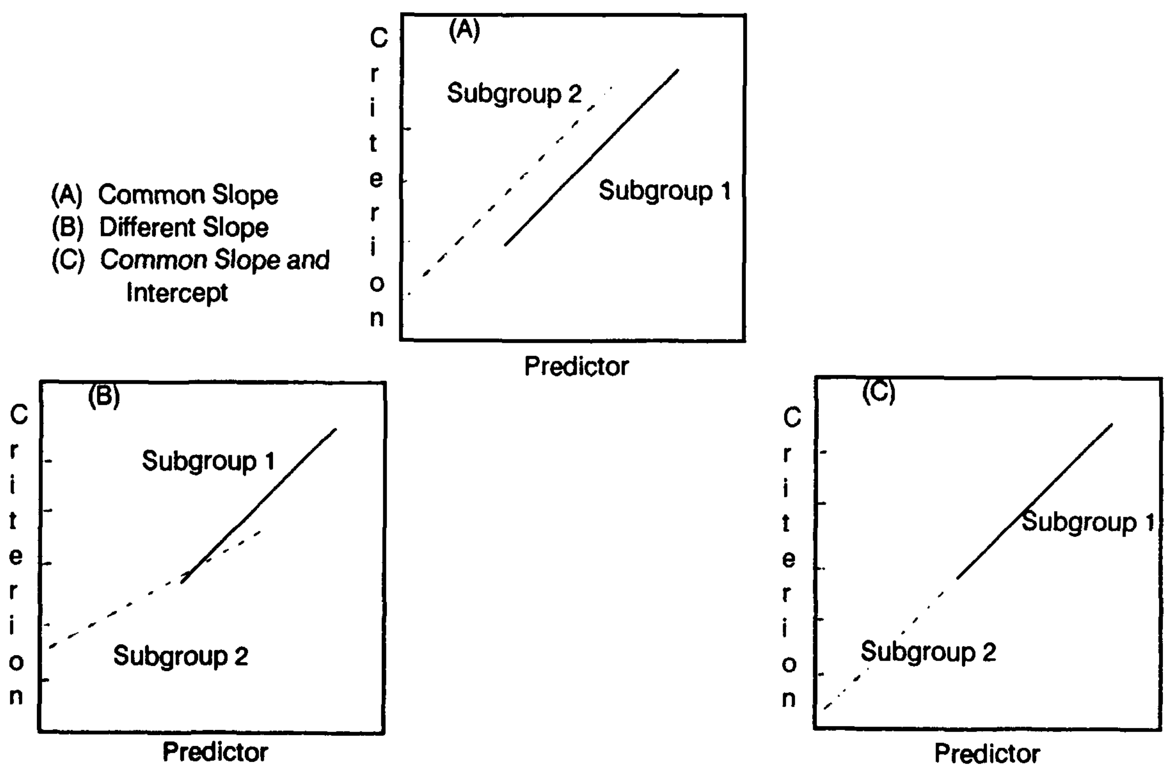

Figure 1. **Schematic Predictor versus Criterion Regression Lines for Two Groups.**

In the situation shown in Figure 1A, the full AFOQT performance of one of the subgroups is consistently overestimated or underestimated, if the common regression line is used (intercept bias). On the other hand, when regression lines are not parallel, another type of bias defined as "slope bias" may be operating; i.e., the differences in predicted AFOQT performance between the subgroups for various QSC raw scores are not constant. Figure 1B illustrates an example of slope bias in which differences in predicted AFOQT scores between Subgroup 1 and Subgroup 2 with the same Quick Score raw score are smaller for the lower range of Quick Score raw scores than for the upper range. If no bias exists, the subgroups will share the same regression line (Figure 1C). It should be noted that the basic assumption of any bias study is that the performance measure--in this case AFOQT composite raw scores--which the Quick Score Composite is designed to predict is not biased itself since accurate prediction of a biased criterion is also discriminatory (Guinn, Tupes, & Alley, 1970).

Tables 15 through 22 summarize the results of the analyses that are presented in detail in Appendix C. The first column shows the functional form of the best linear model for each composite. This describes whether the relationship between the full AFOQT composite and each QSC is linear or curvilinear (Model 1 vs. Model 2). The next column describes whether the groups being tested have the same slope (parallel), different slopes (interaction), or the same slope and intercept (no difference). The third column provides the number of the best linear model (see Table A-3).

Table 15. **Linear Models Analysis Results for Gender Bias: AFOQT Form P1 Point-Biserial QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Curvilinear	Parallel	3
Navigator-Technical	Curvilinear	Parallel	3
Academic Aptitude	Curvilinear	Interaction	1
Verbal	Curvilinear	Interaction	1
Quantitative	Curvilinear	Parallel	3

Table 16. **Linear Models Analysis Results for Gender Bias: AFOQT Form P1 Random QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Parallel	4
Navigator-Technical	Linear	Parallel	4
Academic Aptitude	Linear	No Difference	6
Verbal	Curvilinear	No Difference	5
Quantitative	Linear	Parallel	4

Table 17. **Linear Models Analysis Results for Ethnic Bias: AFOQT Form P1 Point-Biserial QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Interaction	2
Navigator-Technical	Curvilinear	Interaction	1
Academic Aptitude	Linear	Interaction	2
Verbal	Linear	Interaction	2
Quantitative	Curvilinear	Parallel	3

Table 18. **Linear Models Analysis Results for Ethnic Bias: AFOQT Form P1 Random QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Parallel	4
Navigator-Technical	Linear	Parallel	4
Academic Aptitude	Curvilinear	Parallel	3
Verbal	Curvilinear	Parallel	3
Quantitative	Linear	Parallel	4

Table 19. **Linear Models Analysis Results for Gender Bias: AFOQT Form P2 Point-Biserial QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Parallel	4
Navigator-Technical	Curvilinear	Parallel	3
Academic Aptitude	Curvilinear	Parallel	3
Verbal	Curvilinear	Parallel	3
Quantitative	Linear	Parallel	4

Table 20. **Linear Models Analysis Results for Gender Bias: AFOQT Form P2 Random QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Curvilinear	No Difference	5
Navigator-Technical	Linear	No Difference	6
Academic Aptitude	Linear	No Difference	6
Verbal	Linear	No Difference	6
Quantitative	Linear	No Difference	6

Table 21. **Linear Models Analysis Results for Ethnic Bias: AFOQT Form P2 Point-Biserial QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Interaction	2
Navigator-Technical	Linear	Parallel	4
Academic Aptitude	Linear	Interaction	2
Verbal	Linear	Interaction	2
Quantitative	Linear	Parallel	4

Table 22. **Linear Models Analysis Results for Ethnic Bias: AFOQT Form P2 Random QSCs**

AFOQT composite	Functional form	Groups	Best model
Pilot	Linear	Parallel	4
Navigator-Technical	Linear	Parallel	4
Academic Aptitude	Linear	Parallel	4
Verbal	Linear	Parallel	4
Quantitative	Curvilinear	Parallel	3

Results indicated that both item selection methods introduced some gender and ethnic bias into Quick Score Composites. All ten point-biserial QSCs showed either slope or intercept gender bias, compared to only three for the random method. Eight of those point-biserial QSCs showed intercept bias; two QSCs exhibited slope bias. Only intercept bias was shown by the random QSCs.

Whether developed randomly or psychometrically, all QSCs exhibited some form of ethnic bias. Seven of the point-biserial QSCs evidenced slope bias, thus producing an inconsistency of overprediction and underprediction across the range of predicted scores. Randomly developed QSCs also exhibited error in prediction with the introduction of intercept bias. However, intercept bias produces a consistent error of prediction across the range of scores.

Although both methods showed some evidence of bias, the effect was less pronounced overall for random QSCs. The difference in random-QSC-predicted scores between gender groups was found to be minimal (approximately 3 raw score points) and the magnitude of difference between ethnic groups was also not appreciable. For those random QSCs where bias did exist, the minority groups were favored. Therefore, no efforts were made to apply statistical adjustments to the resulting tables for specific gender or ethnic subgroups. Overall, the random item selection method evidenced fewer instances of potential bias when compared to the point-biserial method.

IV. CONCLUSIONS

Results of the analyses performed, along with the operational needs of the Air Force Recruiting Service, must be considered in recommending the best item selection method.

Given the criteria for this study--optimal prediction of success on the AFOQT Form P, reliability, and minimal gender and ethnic effects--results favored within-subtest random item selection. The forced selection of roughly 27% of total items from each AFOQT subtest controlled the overrepresentation of items from certain subtests and allowed equal taxonomic representation of all subtests in their respective Quick Score Composites.

Reliabilities for random and point-biserial QSCs were comparable, with the point-biserial QSCs being slightly more reliable. However, though screening scores must be reliable, it is equally important that they demonstrate high validity. In fact, both the random and point-biserial item selection methods produced strong significant positive correlations with AFOQT Forms P1 and P2. However, the question of fairness--not investigated by Rogers (1985), Valentine (1961), Miller (1966), or Miller (1968)--must also play a major role in the selection of a method. Unfairness in either the random or point-biserial QSCs could lead to problems for the prescreening system.

The 1- to 2-week delay between testing and availability of AFOQT composite scores to the recruiter is crucial. The recruiter must maintain motivation on the part of the potentially qualified applicants, and should concentrate his/her attention on these applicants. If the QSC is underpredicting the AFOQT composites for some particular population group, the recruiter may end up losing underpredicted prospects from that group

because of inadequate attention to them. Thus, it is important that the QSCs demonstrate equal precision for all population subgroups. Results of this analysis indicate the common regressions (i.e., those which disregard gender or ethnicity) for some of the random QSCs and all of the point-biserial QSCs result in inaccurate prediction of AFOQT composite raw scores for some subgroup. The random QSCs were favored in this regard because they evidenced fewer instances of bias and because in those cases where bias was found, the common regressions favored the minority groups to a small degree.

Further, this study's findings are at variance with those of Goh (1979), who reported that psychometric item selection methods are more desirable and meaningful than the systematic random selection of items. However, Goh used the Yudin (1966) "random" selection method, which may not be truly random due to the fact that it selects items from subtests on the basis of every Xth item, and allows X to vary from one subtest to another. When a true random item selection method is applied, more desirable and meaningful short-form composites are developed.

Random item selection--by its mathematical nature--produces equal item-taxonomic representation. It also produced high reliability, and high positive correlations with the full-length scales; and it introduced less gender or ethnic bias than that associated with the other method. QSC scores based on within-subtest random item selection were effective predictors of the full-length scales.

REFERENCES

Air Force Regulation 35-8. (1983). *Air Force Military Personnel Testing System.* Washington, DC: Department of the Air Force.

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory.* California: Wadsworth.

Berger, F.R., Gupta, W.M., Berger, R.M., & Skinner, J. (1988). *Air Force Officer Qualifying Test (AFOQT) Form P: Test construction* (AFHRL-TR-88-30, AD-A200 678). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Bryson, R. (1972). Shortening tests: Effects of method used, length, and internal consistency on correlation with total score. *Proceedings of the 80th Annual American Psychological Association Convention.* Honolulu, Hawaii.

Cole, N.S. (1973). Bias in selection. *Journal of Educational Measurement, 10,* 237-255.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Goh, D.S. (1979). Empirical versus random item selection in the design of intelligence test short forms - the WISC-R example. *Applied Psychological Measurement, 3*(1), 75-82.

Guinn, N., Tupes, E.C., & Alley, W.E. (1970). *Cultural subgroup differences in the relationships between Air Force aptitude composites and training criteria* (AFHRL-TR-70-35, AD-715 922). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division.

Henrysson, S. (1971). Item Analyses. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 130-159). Washington, DC: American Council on Education.

Lautenschlager, G.J., & Mendoza, J.L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133-139.

Levy, P. (1968). Short-form tests: A methodological review. *Psychological Bulletin*, *69*(6) 410-416.

Miller, R.E. (1966). *Development of officer selection and classification tests-1966* (PRL-TR-66-5, AD-639 237). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Miller, R.E. (1968). *Development of officer selection and classification tests-1968* (AFHRL-TR-68-104, AD-679 989). Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division.

Mosier, C.I. (1951). Symposium: The need and means of cross-validation I. Problems and designs of cross-validation. *Educational and Psychological Measurement*, *11*, 5-11.

Mumpower, D.L. (1964). The fallacy of the short form. *Journal of Clinical Psychology*, *20*, 111-113.

Neter, J., Wasserman, W., & Whitmore, G.A. (1988). *Applied Statistics* (3rd ed.). Newton, MA: Allyn and Bacon.

Rogers, D.L. (1983). Development of the Air Force officer screening composites. *Proceedings of the 25th Annual Conference of the Military Testing Association* (pp. 467-471). Gulf Shores, AL.

Rogers, D.L. (1985). *Screening composites for Air Force officers* (AFHRL-TP-85-2, AD-A154 315). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Rogers, D.L., Roach, B.W., & Short, L.O. (1986). *Mental ability testing in the selection of Air Force officers: A brief historical overview* (AFHRL-TP-86-23, AD-A173 484). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Rogers, D.L., Roach, B.W., & Wegner, T.G. (1986). *Air Force Officer Qualifying Test Form O: Development and standardization* (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Steuck, K.W., Watson, T.W., Skinner, M.J. (1988). Air Force Officer Qualifying Test AFOQT: Forms P pre-implementation analyses and equating (AFHRL-TR-88-6, AD-A201 100). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Valentine, L.D., Jr. (1961). *Development of the Air Force Precommission Screening Test-62* (ASD-TN-61-146, AD-269 527). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Valentine, L.D., Jr., & Creager, J.A. (1961). *Officer selection and classification tests: Their development and use* (ASD-TN-61-145). Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division.

Ward, J.H., Jr., & Jennings, E. (1973). *Introduction to linear models*. New Jersey: Prentice Hall.

Wegner, T.G., & Short, L.O. (1986). Assessing the accuracy of the AFOQT quick score procedure. *Proceedings of the 28th Annual Conference of the Military Testing Association* (pp. 60-65). Mystic, CT.

Wherry, R.J., & Gaylord, R.H. (1943). The concept of test and item reliability in relation to factor pattern. *Psychometrika*, *8*(4), 247-264.

Yudin, L.W. (1966). An abbreviated form of the WISC for use with emotionally disturbed children. *Journal of Consulting Psychology*, *30*, 272-275.

APPENDIX A: LINEAR MODEL SPECIFICATIONS FOR GENDER AND ETHNIC EFFECTS

Table A-1. **Linear Model Specifications for Analysis of Gender Effects**

Model	Components
1	$Y' = U + M + F + QSC_M + QSC_F + QSC_M^2 + QSC_F^2$
2	$Y' = U + M + F + QSC_M + QSC_F$
3	$Y' = U + M + F + QSC + QSC^2$
4	$Y' = U + M + F + QSC$
5	$Y' = U + QSC + QSC^2$
6	$Y' = U + QSC$

Note. These six models were computed for each of the five AFOQT composites.

Y' = predicted full AFOQT composite raw score.
U = unit vector.
M = 1 if male; 0 otherwise.
F = 1 if female; 0 otherwise
QSC_M = Quick Score raw score if male; 0 otherwise.
QSC_F = Quick Score raw score if female; 0 otherwise.
QSC_M^2 = Quick Score raw score squared if male; 0 otherwise.
QSC_F^2 = Quick Score raw score squared if female; 0 otherwise.

Table A-2. **Linear Model Specifications for Analysis of Ethnicity Effects**

Model	Components
1	$Y' = U + C + B + H + A + QSC_C + QSC_B + QSC_H + QSC_A + QSC_C^2 + QSC_B^2 + QSC_H^2 + QSC_A^2$
2	$Y' = U + C + B + H + A + QSC_C + QSC_B + QSC_H + QSC_A$
3	$Y' = U + C + B + H + A + QSC + QSC^2$
4	$Y' = U + C + B + H + A + QSC$
5	$Y' = U + QSC + QSC^2$
6	$Y' = U + QSC$

Note. These six models were computed for each of the five AFOQT composites.

Y' = predicted full AFOQT composite raw score.
U = unit vector.
C = 1 if Caucasian; 0 otherwise.
B = 1 if Black; 0 otherwise.
H = 1 if Hispanic; 0 otherwise.
A = 1 if Asian; 0 otherwise.
QSC_C = Quick Score raw score if Caucasian; 0 otherwise.
QSC_B = Quick Score raw score if Black; 0 otherwise.
QSC_H = Quick Score raw score if Hispanic; 0 otherwise.
QSC_A = Quick Score raw score if Asian; 0 otherwise.
QSC_C^2 = Quick Score raw score squared if Caucasian; 0 otherwise.
QSC_B^2 = Quick Score raw score squared if Black; 0 otherwise.
QSC_H^2 = Quick Score raw score squared if Hispanic; 0 otherwise.
QSC_A^2 = Quick Score raw score squared if Asian; 0 otherwise.

Table A-3. **Summary of the Functional Form Specified by Each Model and the Between-Group Relationship**

Model	Functional form	Between-group relationship
1	Curvilinear	Interaction
2	Linear	Interaction
3	Curvilinear	Parallel
4	Linear	Parallel
5	Curvilinear	No Difference
6	Linear	No Difference

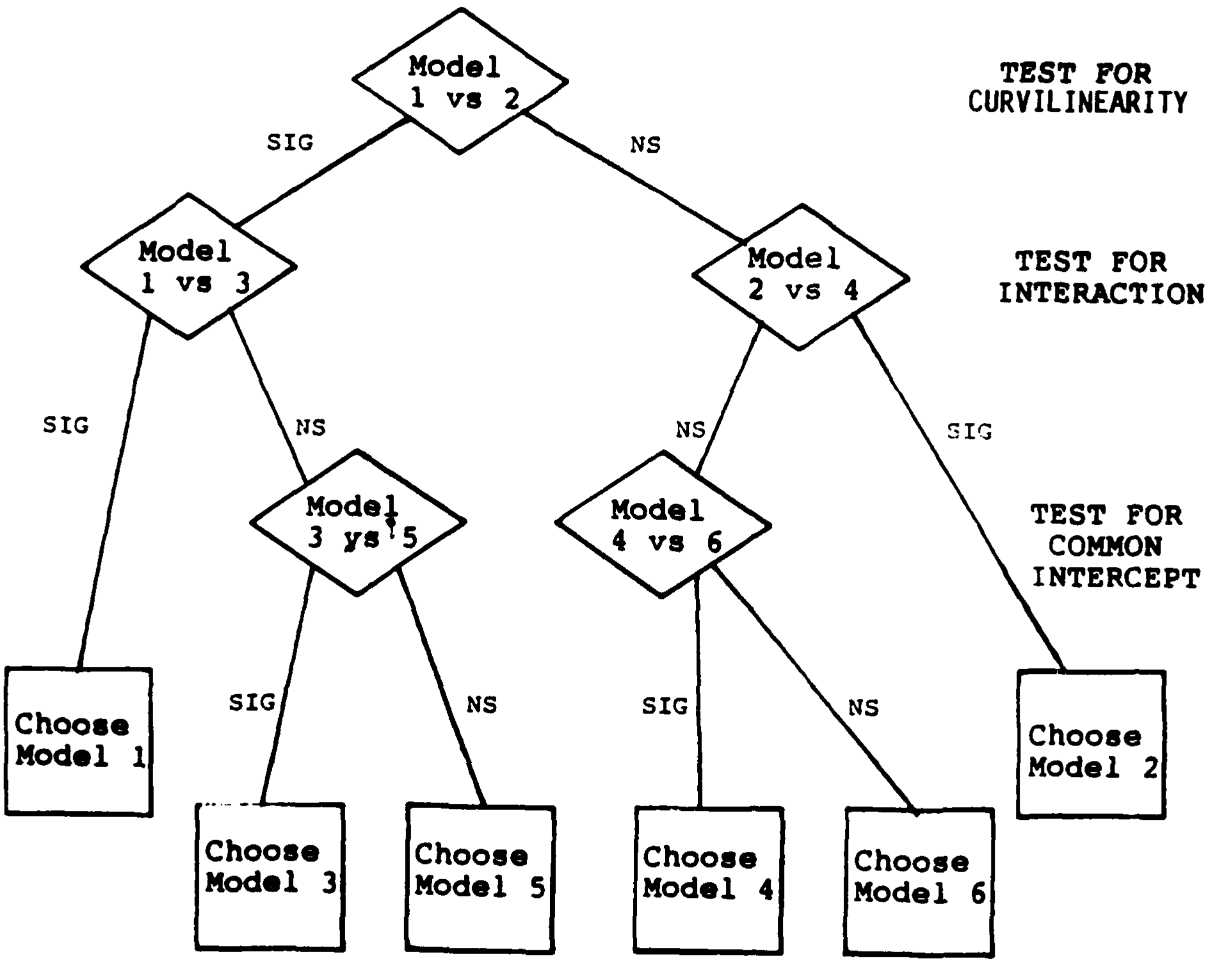

Figure A-1. **Sequential F-test Comparisons.**

APPENDIX B: AFOQT FORMS P1 AND P2 REPLICATION AND CROSS-VALIDATION ANALYSES RESULTS

Table B-1. Mean Composite Difficulty Levels of Form P1 for Full, Pbis, and Random Replication Samples

Composite	AFOQT Form P1	Pbis QSC	Diff	Random QSC	Diff
Pilot	.58	.53	.05	.61	-.03
Navigator-Technical	.59	.53	.06	.63	-.04
Academic Aptitude	.63	.58	.05	.66	-.03
Verbal	.62	.60	.02	.66	-.04
Quantitative	.62	.56	.06	.66	-.04

Table B-2. Mean Composite Difficulty Levels of Form P2 for Full, Pbis, and Random Replication Samples

Composite	AFOQT Form P1	Pbis QSC	Random Diff	QSC	Diff
Pilot	.57	.52	.05	.59	-.02
Navigator-Technical	.58	.52	.06	.60	-.02
Academic Aptitude	.64	.57	.07	.63	.01
Verbal	.65	.58	.07	.65	.00
Quantitative	.63	.56	.07	.61	.02

Table B-3. Summary Statistics of AFOQT Form P1 Point-Biserial QSCs Replication Samples

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	28.74	11.08	-.79	-.13	1	52
Navigator-Technical	37.55	14.89	-.87	-.07	4	70
Academic Aptitude	24.31	9.28	-.78	-.22	0	42
Verbal	12.60	4.91	-.72	-.28	0	21
Quantitative	11.71	5.55	-1.09	-.10	0	21

Table B-4. Summary Statistics of AFOQT Form P1 Random QSCs Replication Samples

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	32.83	7.70	-.27	-.31	8	52
Navigator-Technical	44.82	10.84	-.40	-.33	11	69
Academic Aptitude	27.84	7.24	-.36	-.47	5	42
Verbal	13.93	3.75	-.20	-.51	1	21
Quantitative	11.92	4.48	-.81	-.33	2	21

Table B-5. **Summary Statistics of AFOQT Form P2 Point-Biserial QSC Replication Samples**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	28.27	10.58	-.71	-.16	1	53
Navigator-Technical	36.61	14.16	-.79	-.10	2	68
Academic Aptitude	23.93	9.80	-.90	-.17	1	42
Verbal	12.14	5.34	-.96	-.15	0	21
Quantitative	11.79	5.70	-1.10	-.12	0	21

Table B-6. **Summary Statistics AFOQT Form P2 Random QSCs Replication Samples**

Composite	Mean	SD	Kurtosis	Skew	Min	Max
Pilot	32.03	8.02	-.44	-.23	6	51
Navigator-Technical	42.36	10.44	-.51	-.17	12	68
Academic Aptitude	26.47	7.72	-.73	-.24	5	42
Verbal	13.56	4.35	-.74	-.29	2	21
Quantitative	12.91	4.44	-.81	-.16	1	21

Table B-7. **Coefficient of Variation of AFOQT Form P1 Full, Pbis, and Random Replication Composites**

Composite	AFOQT	Pbis QSC	Random QSC
Pilot	24.01	38.55	23.45
Navigator-Technical	24.40	39.65	24.19
Academic Aptitude	26.87	38.17	26.01
Verbal	29.03	38.97	26.92
Quantitative	30.28	47.40	32.18

Table B-8. **Coefficient of Variation of AFOQT Form P2 Full, Pbis, and Random Replication Composites**

Composite	AFOQT	Pbis QSC	Random QSC
Pilot	23.61	37.42	25.04
Navigator-Technical	24.31	38.68	24.65
Academic Aptitude	27.26	40.95	29.17
Verbal	28.71	43.99	32.08
Quantitative	31.53	48.35	34.39

Table B-9. **AFOQT Form P1 Reliabilities for Full, Random and Pbis Replication Samples**

Composite	Reliability		
	Full	Pbis	Random
Pilot	.958	.931	.863
Navigator-Technical	.971	.949	.911
Academic Aptitude	.961	.921	.883
Verbal	.931	.854	.781
Quantitative	.941	.893	.846

Table B-10. **AFOQT Form P2 Reliabilities for Full, Random and Pbis Replication Samples**

Composite	Reliability		
	Full	Pbis	Random
Pilot	.957	.925	.868
Navigator-Technical	.969	.942	.901
Academic Aptitude	.963	.929	.894
Verbal	.937	.880	.835
Quantitative	.943	.896	.832

Table B-11. **Cross-Validation of Correlations Between the Full AFOQT and Corresponding Pbis and Random QSCs on Form P1**

	Pbis			Random		
	Validation	C-V	Diff[a]	Validation	C-V	Diff[a]
Pilot	.940	.938	.002	.942	.945	-.003
Navigator-Technical	.955	.954	.001	.964	.963	.001
Academic Aptitude	.957	.956	.001	.950	.951	-.001
Verbal	.942	.939	.003	.905	.903	.002
Quantitative	.938	.937	.001	.941	.936	.005

[a]Difference between Validation and C-V correlations.

Table B-12. **Cross-Validation of Correlations Between the Full AFOQT and Corresponding Pbis and Random QSCs on Form P2**

	Pbis			Random		
	Validation	C-V	Diff[a]	Validation	C-V	Diff[a]
Pilot	.937	.931	.006	.956	.951	.005
Navigator-Technical	.956	.952	.004	.964	.959	.005
Academic Aptitude	.960	.957	.003	.958	.955	.003
Verbal	.939	.937	.002	.934	.932	.002
Quantitative	.948	.944	.004	.935	.934	.001

[a]Difference between Validation and C-V correlations.

Table B-13. R^2 Results of the Cross-Validation of the Full AFOQT and Corresponding Pbis QSCs on Form P1

	R^2s		
	Validation	Cross-Validation	Difference[a]
Pilot	.884495	.880095	.004400
Navigator-Technical	.913381	.910879	.002502
Academic Aptitude	.917555	.915617	.001938
Verbal	.887554	.882745	.004809
Quantitative	.880567	.879641	.009260

[a]Difference between R^2s significant at $p < .01$.

Table B-14. R^2 Results of Cross-Validation of the Full AFOQT and Corresponding Random QSCs on Form P1

	R^2s		
	Validation	Cross-Validation	Difference[a]
Pilot	.888062	.893115	-.005053
Navigator-Technical	.930390	.928958	.001432
Academic Aptitude	.904267	.905183	-.000916
Verbal	.820015	.816084	.003931
Quantitative	.887130	.877719	.009411

[a]Difference between R^2s significant at $p < .01$.

Table B-15. R^2 Results of the Cross-Validation of the Full AFOQT and Corresponding Pbis QSCs on Form P2

	R^2s		
	Validation	Cross-Validation	Difference[a]
Pilot	.878221	.866763	.011458
Navigator-Technical	.915703	.906958	.008745
Academic Aptitude	.921693	.916367	.005326
Verbal	.881739	.878831	.002908
Quantitative	.899584	.891361	.008223

[a]Difference between R^2s significant at $p < .01$.

Table B-16. R^2 Results of Cross-Validation of the Full AFOQT and Corresponding Random QSCs on Form P2

	R^2s		
	Validation	Cross-Validation	Difference[a]
Pilot	.915128	.905183	.009945
Navigator-Technical	.929943	.920045	.009898
Academic Aptitude	.919215	.913564	.005651
Verbal	.874210	.870042	.004168
Quantitative	.874393	.872849	.001544

[a]Difference between R^2s significant at $p < .01$.

APPENDIX C: REGRESSION ANALYSES OF AFOQT FORM P1 AND P2 TO DETERMINE GENDER AND ETHNIC BIAS

Table C-1. AFOQT Form P1 Regression Analyses to Determine Gender Bias

Model comparison Full	Model comparison Rest[a]	R^2 Full	R^2 Rest[a]	df_1	df_2	F	Std Err Est Full	Std Err Est Rest[a]
				QSC-P Point-Biserial				
1	2	.8898	.8889	2	1581	6.82*	9.38	9.41
1	3	.8898	.8896	2	1581	1.08	9.38	9.38
3	5	.8896	.8857	1	1583	55.45*	9.38	9.54
				QSC-P Random				
1	2	.8901	.8900	2	1581	.51	9.37	9.37
2	4	.8900	.8900	1	1583	.55	9.37	9.36
4	6	.8900	.8882	1	1584	25.87*	9.36	9.44
				QSC-NT Point-Biserial				
1	2	.9186	.9165	2	1581	20.51*	10.95	11.08
1	3	.9186	.9184	2	1581	1.53	10.95	10.95
3	5	.9184	.9157	1	1583	53.57*	10.95	11.13
				QSC-NT Random				
1	2	.9315	.9314	2	1581	.61	10.05	10.04
2	4	.9314	.9314	1	1583	.99	10.04	10.04
4	6	.9314	.9306	1	1584	17.87*	10.04	10.10
				QSC-AA Point-Biserial				
1	2	.9190	.9183	2	1581	6.81*	7.28	7.31
1	3	.9190	.9184	2	1581	5.77*	7.28	7.30
				QSC-AA Random				
1	2	.9046	.9044	2	1581	2.07	7.90	7.91
2	4	.9044	.9043	1	1583	.54	7.91	7.90
4	6	.9043	.9043	1	1584	1.18	7.90	7.90
				QSC-V Point-Biserial				
1	2	.8891	.8885	2	1581	4.78*	4.54	4.56
1	3	.8891	.8884	2	1581	4.99*	4.54	4.56
				QSC-V Random				
1	2	.8231	.8217	2	1581	6.69*	5.74	5.76
1	3	.8231	.8226	2	1581	2.62	5.74	5.75
3	5	.8226	.8220	1	1583	5.76	5.75	5.75
				QSC-Q Point-Biserial				
1	2	.8846	.8826	2	1581	13.99*	4.97	5.01
1	3	.8846	.8843	2	1581	1.80	4.97	4.98
3	5	.8843	.8831	1	1583	16.80*	4.98	5.00
				QSC-Q Random				
1	2	.8875	.8874	2	1581	.31	4.91	4.91
2	4	.8874	.8874	1	1583	.09	4.91	4.91
4	6	.8874	.8867	1	1584	10.49*	4.91	4.9

N = 1,587.

[a]Restricted model.

*Significant at $p < .01$.

Table C-2. **AFOQT Form P1 Regression Analyses to Determine Ethnicity Bias**

Model comparison		R^2					Std Err Est	
Full	Rest[a]	Full	Rest[a]	df_1	df_2	F	Full	Rest[a]
				QSC-P Point-Biserial				
1	2	.8909	.8908	4	1575	.27	9.35	9.34
2	4	.8908	.8898	3	1579	4.69*	9.34	9.38
				QSC-P Random				
1	2	.8926	.8925	4	1575	.58	9.28	9.27
2	4	.8925	.8922	3	1579	1.02	9.27	9.27
4	6	.8922	.8882	3	1582	19.85*	9.27	9.44
				QSC-NT Point-Biserial				
1	2	.9191	.9182	4	1575	4.23*	10.94	10.98
1	3	.9191	.9178	6	1575	4.10*	10.94	11.00
				QSC-NT Random				
1	2	.9326	.9324	4	1575	.87	9.98	9.98
2	4	.9324	.9323	3	1579	1.18	9.98	9.98
4	6	.9323	.9306	3	1582	13.31*	9.98	10.10
				QSC-AA Point-Biserial				
1	2	.9213	.9212	4	1575	.52	7.19	7.19
2	4	.9212	.9199	3	1579	8.36*	7.19	7.24
				QSC-AA Random				
1	2	.9061	.9053	4	1575	3.42*	7.85	7.88
1	3	.9061	.9056	6	1575	1.26	7.85	7.86
3	5	.9056	.9045	3	1581	6.10*	7.86	7.90
				QSC-V Point-Biserial				
1	2	.8823	.8916	4	1575	2.87	4.49	4.50
2	4	.8916	.8906	3	1579	5.32*	4.50	4.52
				QSC-V Random				
1	2	.8259	.8233	4	1575	5.83*	5.71	5.74
1	3	.8259	.8245	6	1575	2.09	5.71	5.72
3	5	.8245	.8220	3	1581	7.69*	5.72	5.75
				QSC-Q Point-Biserial				
1	2	.8859	.8847	4	1575	4.11	4.95	4.97
1	3	.8859	.8855	6	1575	.86	4.95	4.95
3	5	.8855	.8831	3	1581	11.12*	4.95	5.00
				QSC-Q Random				
1	2	.8900	.8892	4	1575	2.76	4.86	4.88
2	4	.8892	.8892	3	1579	.25	4.88	4.87
4	6	.8892	.8867	3	1582	11.81*	4.87	4.92

N = 1,587.

[a]Restricted model.

*Significant at $p < .01$.

Table C-3. **AFOQT Form P2 Regression Analyses to Determine Gender Bias**

Model comparison		R^2					Std Err Est	
Full	Rest[a]	Full	Rest[a]	df_1	df_2	F	Full	Rest[a]
				QSC-P Point-Biserial				
1	2	.8847	.8841	2	1447	3.70	9.67	9.69
2	4	.8841	.8839	1	1449	2.35	9.69	9.69
4	6	.8839	.8785	1	1450	68.38*	9.69	9.91
				QSC-P Random				
1	2	.9162	.9155	2	1447	6.22*	8.24	8.27
1	3	.9162	.9162	2	1447	.60	8.24	8.24
3	5	.9162	.9160	1	1449	2.58	8.24	8.24
				QSC-NT Point-Biserial				
1	2	.9202	.9192	2	1447	8.80*	10.85	10.90
1	3	.9202	.9200	2	1447	1.76	10.85	10.85
3	5	.9200	.9174	1	1449	45.99*	10.85	11.02
				QSC-NT Random				
1	2	.9303	.9302	2	1447	1.22	10.14	10.14
2	4	.9302	.9301	1	1449	.33	10.14	10.14
4	6	.9301	.9301	1	1450	.12	10.14	10.13
				QSC-AA Point-Biserial				
1	2	.9243	.9229	2	1447	13.46*	7.28	7.34
1	3	.9243	.9241	2	1447	1.89	7.28	7.28
3	5	.9241	.9240	1	1449	3.19	7.28	7.29
				QSC-AA Random				
1	2	.9187	.9185	2	1447	1.68	7.54	7.55
2	4	.9185	.9184	1	1449	1.36	7.55	7.55
4	6	.9184	.9184	1	1450	.28	7.55	7.55
				QSC-V Point-Biserial				
1	2	.8832	.8818	2	1447	9.00*	4.82	4.84
1	3	.8832	.8828	2	1447	2.82	4.82	4.82
3	5	.8828	.8828	1	1449	.02	4.82	4.82
				QSC-V Random				
1	2	.8732	.8730	2	1447	1.38	5.02	5.02
2	4	.8730	.8728	1	1449	1.88	5.02	5.02
4	6	.8728	.8728	1	1450	.06	5.02	5.02
				QSC-Q Point-Biserial				
1	2	.9016	.9014	2	1447	1.82	4.70	4.70
2	4	.9014	.9012	1	1449	3.19	4.70	4.70
4	6	.9012	.9003	1	1450	13.22*	4.70	4.72
				QSC-Q Random				
1	2	.8751	.8749	2	1447	.71	5.29	5.29
2	4	.8749	.8749	1	1449	.27	5.29	5.29
4	6	.8749	.8746	1	1450	3.93	5.29	5.30

N = 1,587.

[a]Restricted model.

*Significant at $p < .01$.

Table C-4. AFOQT Form P2 Regression Analyses to Determine Ethnicity Bias

Model comparison Full	Rest[a]	R^2 Full	R^2 Rest[a]	df_1	df_2	F	Std Err Est Full	Std Err Est Rest[a]
				QSC-P Point-Biserial				
1	2	.8906	.8902	4	1441	1.38	9.44	9.44
2	4	.8902	.8892	3	1445	4.24	9.44	9.48
				QSC-P Random				
1	2	.9192	.9191	4	1441	2.18	8.09	8.11
2	4	.9191	.9187	3	1445	2.14	8.11	8.11
4	6	.9187	.9151	3	1448	21.84*	8.11	8.29
				QSC-NT Point-Biserial				
1	2	.9218	.9216	4	1441	1.05	10.76	10.76
2	4	.9216	.9211	3	1445	3.22	10.76	10.78
4	6	.9211	.9160	3	1448	30.97*	10.78	11.11
				QSC-NT Random				
1	2	.9324	.9321	4	1441	1.37	10.00	10.01
2	4	.9321	.9318	3	1445	2.15	10.01	10.02
4	6	.9318	.9301	3	1448	11.82*	10.02	10.13
				QSC-AA Point-Biserial				
1	2	.9270	.9267	4	1441	1.59	7.16	7.17
2	4	.9267	.9241	3	1445	7.32*	7.17	7.22
				QSC-AA Random				
1	2	.9221	.9218	4	1441	1.36	7.40	7.40
2	4	.9218	.9217	3	1445	.99	7.40	7.40
4	6	.9217	.9184	3	1448	20.01*	7.40	7.55
				QSC-V Point-Biserial				
1	2	.8884	.8881	4	1441	.91	4.72	4.72
2	4	.8881	.8865	3	1445	7.00*	4.72	4.75
				QSC-V Random				
1	2	.8822	.8817	4	1441	1.56	4.85	4.85
2	4	.8817	.8810	3	1445	2.80	4.85	4.86
4	6	.8810	.8728	3	1448	33.08*	4.86	5.02
				QSC-Q Point-Biserial				
1	2	.9036	.9035	4	1441	.09	4.66	4.65
2	4	.9035	.9032	3	1445	1.75	4.65	4.66
4	6	.9032	.9003	3	1448	14.42*	4.66	4.72
				QSC-Q Random				
1	2	.8793	.8779	4	1441	4.30*	5.21	5.24
1	3	.8793	.8781	6	1441	2.24	5.21	5.23
3	5	.8781	.8746	3	1447	14.08*	5.23	5.30

N = 1,587.

[a]Restricted model.

*Significant at $p < .01$.

APPENDIX D: MEANS AND STANDARD DEVIATIONS OF THE FULL, POINT-BISERIAL, AND RANDOM AFOQT FORM P1 COMPOSITES BY GENDER AND ETHNIC GROUP

Table D-1. AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Male Sample[a]

	Full		Pbis		Random	
Composite	Mean	SD	Mean	SD	Mean	SD
Pilot	121.81	27.31	29.83	10.94	33.53	7.30
Navigator-Technical	162.18	37.14	39.15	14.79	45.92	10.51
Academic Aptitude	96.06	25.08	24.80	9.48	28.33	7.08
Verbal	49.57	13.49	12.62	5.06	14.07	3.70
Quantitative	49.49	14.30	12.18	5.59	14.26	4.41

[a]N = 1,285.

Table D-2. AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Female Sample[a]

	AFOQT		Pbis		Random	
Composite	Mean	SD	Mean	SD	Mean	SD
Pilot	101.62	25.68	23.19	9.96	28.62	6.95
Navigator-Technical	136.68	35.92	30.70	13.77	39.21	10.34
Academic Aptitude	88.15	26.32	21.69	9.46	25.80	7.60
Verbal	45.78	14.09	11.92	5.25	13.53	4.00
Quantitative	42.37	14.51	9.77	5.38	12.27	4.50

[a]N = 320.

Table D-3. AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composites American Indian/Alaskan Native Sample[a]

	AFOQT		Pbis		Random	
Composite	Mean	SD	Mean	SD	Mean	SD
Pilot	111.31	17.75	26.08	8.70	31.00	5.51
Navigator-Technical	143.54	28.87	33.69	12.65	41.46	10.15
Academic Aptitude	81.54	21.49	21.00	9.13	25.39	6.48
Verbal	40.15	9.97	10.00	4.60	13.08	2.69
Quantitative	41.39	14.77	11.00	5.83	12.31	5.11

[a]N = 13.

Table D-4. AFOQT Form P1 Means and Standard Deviations for Full Point-Biserial, and Random Composite Asian/Pacific Islander Sample[a]

	AFOQT		Pbis		Random	
Composite	Mean	SD	Mean	SD	Mean	SD
Pilot	107.90	25.46	26.08	10.24	30.37	7.42
Navigator-Technical	150.26	35.74	35.62	13.72	43.17	10.69
Academic Aptitude	87.93	25.47	22.12	9.27	26.11	6.98
Verbal	40.55	16.00	10.70	5.78	12.11	4.27
Quantitative	47.38	13.03	11.42	5.02	14.00	4.15

[a]N = 78.

Table D-5. **AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Black Sample**[a]

Composite	AFOQT		Pbis		Random	
	Mean	SD	Mean	SD	Mean	SD
Pilot	85.66	24.32	16.79	8.50	24.60	6.73
Navigator-Technical	113.11	33.68	21.46	11.54	32.75	9.83
Academic Aptitude	68.39	25.60	15.03	8.83	20.56	7.50
Verbal	34.90	14.03	8.45	5.06	10.93	4.14
Quantitative	33.49	13.73	6.58	4.79	9.62	4.26

[a]N = 181.

Table D-6. **AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Hispanic Sample**[a]

Composite	AFOQT		Pbis		Random	
	Mean	SD	Mean	SD	Mean	SD
Pilot	100.03	28.72	21.38	10.14	28.27	7.78
Navigator-Technical	133.78	38.15	28.42	13.23	38.08	10.74
Academic Aptitude	74.42	26.75	16.62	9.45	21.78	7.34
Verbal	35.38	15.01	8.57	5.38	10.63	4.08
Quantitative	39.03	14.17	8.05	5.22	11.15	4.40

[a]N = 60.

Table D-7. **AFOQT Form P1 Means and Standard Deviations for Full, Point-Biserial, and Random Composite Caucasian Sample**[a]

Composite	AFOQT		Pbis		Random	
	Mean	SD	Mean	SD	Mean	SD
Pilot	123.84	25.11	30.67	10.25	34.03	6.70
Navigator-Technical	164.96	34.14	40.30	13.92	46.67	9.68
Academic Aptitude	99.61	22.50	25.97	8.70	29.67	6.41
Verbal	48.96	12.05	13.36	4.65	.67	3.31
Quantitative	50.65	13.43	12.61	5.33	14.60	4.17

[a]N = 1,273.

APPENDIX E: MEANS AND STANDARD DEVIATIONS OF THE FULL, POINT-BISERIAL, AND RANDOM AFOQT FORM P2 COMPOSITES BY GENDER AND ETHNIC GROUP

Table E-1. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Male Sample[a]**

	Full		Pbis		Random	
Composite	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pilot	122.46	26.87	30.45	10.59	33.64	7.84
Navigator-Technical	159.57	36.39	39.08	14.25	44.25	10.17
Academic Aptitude	97.28	25.54	24.83	9.80	27.08	7.52
Verbal	48.96	13.67	12.45	5.30	13.71	4.25
Quantitative	48.32	1[illegible]53	12.38	5.65	13.37	4.31

[a]N = 1,201.

Table E-2. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Female Sample[a]**

	Full		Pbis		Random	
Composite	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pilot	101.51	28.45	24.02	10.71	27.63	8.02
Navigator-Technical	135.19	39.69	31.36	14.61	37.24	11.16
Academic Aptitude	89.41	29.02	22.14	10.69	24.74	8.43
Verbal	47.33	15.90	11.80	5.90	13.14	4.82
Quantitative	42.08	15.51	10.33	5.81	11.61	4.61

[a]N = 208.

Table E-3. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites American Indian/Alaskan Native Sample[a]**

	AFOQT		Pbis		Random	
Composite	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pilot	113.09	26.56	27.78	9.38	30.78	7.30
Navigator-Technical	146.00	35.43	33.83	13.33	40.30	10.07
Academic Aptitude	84.26	26.64	20.87	9.63	23.09	8.45
Verbal	41.39	16.84	10.65	5.78	11.57	5.56
Quantitative	42.87	13.93	10.22	5.63	11.52	4.06

[a]N = 23.

Table E-4. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Asian/Pacific Islander Sample[a]**

	AFOQT		Pbis		Random	
Composite	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Pilot	111.13	29.30	27.44	10.19	29.26	7.93
Navigator-Technical	153.61	41.64	37.26	15.44	42.19	11.09
Academic Aptitude	93.13	29.07	23.74	10.88	25.43	8.40
Verbal	44.02	15.84	10.93	5.73	11.96	4.63
Quantitative	49.11	15.98	12.82	6.17	13.46	4.66

[a]N = 54.

Table E-5. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Black Sample[a]**

Composite	AFOQT Mean	AFOQT SD	Pbis Mean	Pbis SD	Random Mean	Random SD
Pilot	78.05	22.86	15.40	8.35	21.40	6.72
Navigator-Technical	102.28	31.41	19.17	11.33	28.66	9.06
Academic Aptitude	62.52	21.95	12.48	7.43	17.55	6.33
Verbal	32.75	12.13	6.72	4.04	9.38	3.68
Quantitative	29.77	12.34	5.76	4.47	8.19	3.76

[a]N = 166.

Table E-6. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Hispanic Sample[a]**

Composite	AFOQT Mean	AFOQT SD	Pbis Mean	Pbis SD	Random Mean	Random SD
Pilot	104.43	28.28	24.33	11.71	28.92	8.20
Navigator-Technical	136.10	37.73	31.16	14.76	38.00	10.66
Academic Aptitude	83.69	28.70	20.47	10.54	23.51	8.68
Verbal	43.55	15.12	10.82	5.60	12.43	4.72
Quantitative	40.14	15.04	9.65	5.63	11.08	4.61

[a]N = 49.

Table E-7. **AFOQT Form P2 Means and Standard Deviations for Full, Point-Biserial, and Random Composites Caucasian Sample[a]**

Composite	AFOQT Mean	AFOQT SD	Pbis Mean	Pbis SD	Random Mean	Random SD
Pilot	124.97	23.88	31.39	9.74	34.33	7.08
Navigator-Technical	163.06	32.65	40.45	13.04	45.13	9.28
Academic Aptitude	101.13	23.05	26.18	9.10	28.11	6.91
Verbal	51.37	12.56	13.24	5.06	14.33	4.03
Quantitative	49.76	13.48	12.93	5.31	13.78	4.04

[a]N = 1,190.

* U. S. GOVERNMENT PRINTING OFFICE: 1990--761-051/20076